RNA-interference by single-stranded RNA molecules

Description

The present invention relates to sequence and structural features of single-stranded (ss)RNA molecules required to mediate target-specific nucleic acid modifications by RNA-interference (RNAi), such as target mRNA degradation and/or DNA methylation.

Most eukaryotes possess a cellular defense system protecting their genomes against invading foreign genetic elements. Insertion of foreign elements is believed to be generally accompanied by formation of dsRNA that is interpreted by the cell as a signal for unwanted gene activity (e.g. Ahlquist, Science 296 (2002), 1270 - 1273; Fire et al., Nature 391 (1998), 806 - 811). Dicer RNase III rapidly processes dsRNA to small dsRNA fragments of distinct size and structure (e.g. Bernstein et al., Nature 409 (2001), 363 - 366), the small interfering RNAs (siRNAs) (Elbashir et al., Genes & Dev. 15 (2001 b), 188 - 200), which direct the sequence-specific degradation of the single-stranded mRNAs of the invading genes. siRNA duplexes have 2- to 3-nt 3' overhanging ends and contain 5' phosphate and free 3' hydroxyl termini (WO 02/44321). The process of posttranscriptional dsRNA-dependent gene silencing is commonly referred to as RNA interference (RNAi), and in some instances is also linked to transcriptional silencing.

Experimental introduction of siRNA duplexes into mammalian cells is now widely used to disrupt the activity of cellular genes homologous in sequence to the introduced dsRNA. Used as a reverse genetic approach, siRNA-induced gene silencing accelerates linking of gene sequence to biological function. siRNA duplexes are short enough to bypass general dsRNA-induced unspecific effects in vertebrate animal and mammalian

cells. siRNAs may also be expressed intracellularly from introduced expression plasmids or viral vectors providing an alternative to chemical RNA synthesis. Therefore, an understanding of how siRNAs act in mammalian systems is important for refining this gene silencing technology and for producing gene-specific therapeutic agents.

Biochemical studies have begun to unravel the mechanistic details of RNAi. The first cell-free systems were developed using *D. melanogaster* cell or embryo extracts, and were followed by the development of *in vitro* systems from *C. elegans* embryo and mouse embryonal carcinoma cells. While the *D. melanogaster* lysates support the steps of dsRNA processing and sequence-specific mRNA targeting, the latter two systems only recapitulate the first step.

RNAi in *D. melanogaster* extracts is initiated by ATP-dependent processing of long dsRNA to siRNAs by Dicer RNase III (e.g. Bernstein et al., (2001), supra). Thereafter, siRNA duplexes are assembled into a multi-component complex, which guides the sequence-specific recognition of the target mRNA and catalyzes its cleavage (e.g. Elbashir (2001 b), supra). This complex is referred to as RNA-induced silencing complex (RISC) (Hammond et al., Nature 404 (2000), 293 - 296). siRNAs in *D. melanogaster* are predominantly 21- and 22-nt, and when paired in a manner to contain a 2-nt 3' overhanging structure effectively enter RISC (Elbashir et al., EMBO J. 20 (2001 c), 6877 - 6888). Mammalian systems have siRNAs of similar size, and siRNAs of 21- and 22-nt also represent the most effective sizes for silencing genes expressed in mammalian cells (e.g. Elbashir et al., Nature 411 (2001 a), 494 - 498, Elbashir et al., Methods 26 (2002), 199 - 213).

RISC assembled on siRNA duplexes in *D. melanogaster* embryo lysate targets homologous sense as well as antisense single-stranded RNAs for degradation. The cleavage sites for sense and antisense target RNAs are

located in the middle of the region spanned by the siRNA duplex. Importantly, the 5'-end, and not the 3'-end, of the guide siRNA sets the ruler for the position of the target RNA cleavage. Furthermore, a 5' phosphate is required at the target-complementary strand of a siRNA duplex for RISC activity, and ATP is used to maintain the 5' phosphates of the siRNAs (Nykänen et al., Cell 107 (2001), 309 - 321). Synthetic siRNA duplexes with free 5' hydroxyls and 2-nt 3' overhangs are so readily phosphorylated in *D. melanogaster* embryo lysate that the RNAi efficiencies of 5'-phosphorylated and non-phosphorylated siRNAs are not significantly different (Elbashir et al. (2001 c), supra).

Unwinding of the siRNA duplex must occur prior to target RNA recognition. Analysis of ATP requirements revealed that the formation of RISC on siRNA duplexes required ATP in lysates of *D. melanogaster*. Once formed, RISC cleaves the target RNA in the absence of ATP. The need for ATP probably reflects the unwinding step and/or other conformational rearrangements. However, it is currently unknown if the unwound strands of an siRNA duplex remain associated with RISC or whether RISC only contains a single-stranded siRNA.

A component associated with RISC was identified as Argonaute2 from *D. melanogaster Schneider 2* (S2) cells (Hammond et al., Science 293 (2001 a), 1146 - 1150), and is a member of a large family of proteins. The family is referred to as Argonaute or PPD family and is characterized by the presence of a PAZ domain and a C-terminal Piwi domain, both of unknown function (Cerutti et al., Trends Biochem. Sci. (2000), 481 - 482); Schwarz and Zamore, Genes & Dev. 16 (2002), 1025 - 1031). The PAZ domain is also found in Dicer. Because Dicer and Argonaute2 interact in S2 cells, PAZ may function as a protein-protein interaction motif. Possibly, the interaction between Dicer and Argonaute2 facilitates siRNA incorporation into RISC. In *D. melanogaster*, the Argonaute family has five members, most of which were shown to be involved in gene silencing and

development. The mammalian members of the Argonaute family are poorly characterized, and some of them have been implicated in translational control, microRNA processing and development. The biochemical function of Argonaute proteins remains to be established and the development of more biochemical systems is crucial.

Here we report on the analysis of human RISC in extracts prepared from HeLa cells. The reconstitution of RISC and the mRNA targeting step revealed that RISC is a ribonucleoprotein complex that is composed of a single-stranded siRNA. Once RISC is formed the incorporated siRNA can no longer exchange with free siRNAs. Surprisingly, RISC can be reconstituted in HeLa S100 extracts providing single-stranded siRNAs. Introducing 5' phosphorylated single-stranded antisense siRNAs into HeLa cells potently silences an endogenous gene with similar efficiency than duplex siRNA.

The object underlying the present invention is to provide novel agents capable of mediating target-specific RNAi.

The solution of this problems is provided by the use of a single-stranded RNA molecule for the manufacture of an agent for inhibiting the expression of said target transcript. Surprisingly, it was found that single-stranded RNA molecules are capable of inhibiting the expression of target transcripts by RNA-interference (RNAi).

The length of the single-stranded RNA molecules is preferably from 14-50 nt, wherein at least the 14 to 20 5'-most nucleotides are substantially complementary to the target RNA transcript. The RNA oligonucleotides may have a free 5' hydroxyl moiety, or a moiety which is 5' phosphorylated (by means of chemical synthesis or enzymatic reactions) or which is modified by 5'-monophosphate analogues.

The inhibition of target transcript expression may occur in vitro, e.g. in eucaryotic, particularly mammalian cell cultures or cell extracts. On the other hand, the inhibition may also occur in vivo i.e. in eucaryotic, particularly mammalian organisms including human beings.

Preferably, the single-stranded RNA molecule has a length from 15 - 29 nucleotides. The RNA-strand may have a 3´hydroxyl group. In some cases, however, it may be preferable to modify the 3' end to make it resistant against 3' to 5' exonucleases. Tolerated 3'-modifications are for example terminal 2'-deoxy nucleotides, 3' phosphate, 2',3'-cyclic phosphate, C3 (or C6, C7, C12) aminolinker, thiol linkers, carboxyl linkers, non-nucleotidic spacers (C3, C6, C9, C12, abasic, triethylene glycol, hexaethylene glycol), biotin, fluoresceine, etc.

The 5´-terminus comprises an OH group, a phosphate group or an analogue thereof. Preferred 5' phosphate modifications are 5'-monophosphate ($(HO)_2(O)P$-O-5'), 5'-diphosphate ($(HO)_2(O)P$-O-P(HO)(O)-O-5'), 5'-triphosphate ($(HO)_2(O)P$-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-guanosine cap (7-methylated or non-methylated) (7m-G-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-adenosine cap (Appp), and any modified or unmodified nucleotide cap structure (N-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-monothiophosphate (phosphorothioate; $(HO)_2(S)P$-O-5'), 5'-monodithiophosphate (phosphorodithioate; (HO)(HS)(S)P-O-5'), 5'-phosphorothiolate ($(HO)_2(O)P$-S-5'); any additional combination of oxgen/sulfur replaced monophosphate, diphosphate and triphosphates (e.g. 5'-alpha-thiotriphosphate, 5'-gamma-thiotriphosphate, etc.), 5'-phosphoramidates ($(HO)_2(O)P$-NH-5', $(HO)(NH_2)(O)P$-O-5'), 5'-alkylphosphonates (R = alkyl = methyl, ethyl, isopropyl, propyl, etc., e.g. RP(OH)(O)-O-5'-, $(OH)_2(O)P$-5'-CH_2-), 5'-alkyletherphosphonates (R = alkylether = methoxymethyl ($MeOCH_2$-), ethoxymethyl, etc., e.g. RP(OH)(O)-O-5'-).

The sequence of the RNA molecule of the present invention has to have a sufficient identity to a nucleic acid target molecule in order to mediate target-specific RNAi. Thus the single-stranded RNA molecule of the present invention is substantially complementary to the target transcript.

The target RNA cleavage reaction guided by the single-stranded RNA molecules of the present invention is highly sequence-specific. However, no all positions of the RNA molecule contribute equally to target recognition. Mismatches, particularly at the 3´-terminus of the single-stranded RNA molecule, more particularly the residues 3' to the first 20 nt of the single-stranded RNA molecule are tolerated. Especially preferred are single-stranded RNA molecules having at the 5'-terminus at least 15 and preferably at least 20 nucleotides which are completely complementary to a predetermined target transcript or have at only mismatch and optionally up to 35 nucleotides at the 3´-terminus which may contain 1 or several, e.g. 2, 3 or more mismatches.

In order to enhance the stability of the single-stranded RNA molecules, the 3´-ends may be stabilized against degradation, e.g. they may be selected such that they consist of purine nucleotides, particularly adenosine or guanosine nucleotides. Alternatively or additionally, 3´nucleotides may be substituted by modified nucleotide analogues, including backbone modifications of ribose and/or phosphate residues.

In an especially preferred embodiment of the present invention the RNA molecule may contain at least one modified nucleotide analogue. The nucleotide analogues may be located at positions where the target-specific activity, e.g. the RNAi mediating activity is not substantially affected, e.g. in a region at the 5'-end and/or the 3'-end of the RNA molecule. Particularly, the 3´-terminus may be stabilized by incorporating modified nucleotide analogues, such as non-nucleotidic chemical derivatives such as

C3 (or C6, C7, C12) aminolinker, thiol linkers, carboxyl linkers, non-nucleotidic spacers (C3, C6, C9, C12, abasic, triethylene glycol, hexaethylene glycol), biotin, fluoresceine, etc. A further modification, by which the nuclease resistance of the RNA molecule may be increased, is by covalent coupling of inverted nucleotides, e.g. 2'-deoxyribonucleotides or ribonucleotides to the 3'-end of the RNA molecule. A preferred RNA molecule structure comprises: 5'-single-stranded siRNA-3'-O-P(O)(OH)-O-3'-N, wherein N is a nucleotide, e.g. a 2'-deoxyribonucleotide or ribonucleotide, typically an inverted thymidine residue, or an inverted oligonucleotide structure, e.g. containing up to 5 nucleotides.

Preferred nucleotide analogues are selected from sugar- or backbone-modified ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides, containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase such as uridines or cytidines modified at the 5-position, e.g. 5-(2-amino)propyl uridine, 5-bromo uridine; 5-methyl-cytidine; adenosines and guanosines modified at the 8-position, e.g. 8-bromo guanosine; deaza nucleotides, e.g. 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine are suitable. In preferred sugar-modified ribonucleotides the 2´ OH-group is replaced by a group selected from H, OR, R, halo, SH, SR, NH_2, NHR, NR_2 or CN, wherein R is C_1-C_6 alkyl, alkenyl, alkynyl or methoxyethoxy, and halo is F, Cl, Br or I. In preferred backbone-modified ribonucleotides the phosphoester group connecting to adjacent ribonucleotides is replaced by a modified group, e.g. a phosphorothioate, phosphorodithioate, N3'-O5'- and/or N5'-O3' phosphoramidate group. It should be noted that the above modifications may be combined. For example, complementary or non-complementary nucleotides at the 3'-terminus, particularly after at least 15, more particularly after at least 20 5'-terminal nucleotides may be modified without significant loss of activity.

The single-stranded RNA molecule of the invention may be prepared by chemical synthesis. Methods of synthesizing RNA molecules are known in the art.

The single-stranded RNAs can also be prepared by enzymatic transcription from synthetic DNA templates or from DNA plasmids isolated from recombinant bacteria and subsequent 5′-terminal modification. Typically, phage RNA polymerases are used such as T7, T3 or SP6 RNA polymerase.

A further aspect of the present invention relates to a method of mediating RNA interference in a cell or an organism comprising the steps:

(a) contacting the cell or organism with the single-stranded RNA molecule of the invention under conditions wherein target-specific nucleic acid modifications may occur and

(b) mediating a target-specific nucleic acid modification effected by the single-stranded RNA towards a target nucleic acid having a sequence portion substantially complementary to the single-stranded RNA.

Preferably the contacting step (a) comprises introducing the single-stranded RNA molecule into a target cell, e.g. an isolated target cell, e.g. in cell culture, a unicellular microorganism or a target cell or a plurality of target cells within a multicellular organism. More preferably, the introducing step comprises a carrier-mediated delivery, e.g. by liposomal carriers and/or by injection. Further suitable delivery systems include Oligofectamine (Invitrogen) and Transit-TKO siRNA Transfection reagent (Mirus)

The method of the invention may be used for determining the function of a gene in a cell or an organism or even for modulating the function of a gene in a cell or an organism, being capable of mediating RNA interference.

The cell is preferably a eukaryotic cell or a cell line, e.g. a plant cell or an animal cell, such as a mammalian cell, e.g. an embryonic cell, a pluripotent stem cell, a tumor cell, e.g. a teratocarcinoma cell or a virus-infected cell. The organism is preferably a eukaryotic organism, e.g. a plant or an animal, such as a mammal, particularly a human.

The target gene to which the RNA molecule of the invention is directed may be associated with a pathological condition. For example, the gene may be a pathogen-associated gene, e.g. a viral gene, a tumor-associated gene or an autoimmune disease-associated gene. The target gene may also be a heterologous gene expressed in a recombinant cell or a genetically altered organism. By determinating or modulating, particularly, inhibiting the function of such a gene valuable information and therapeutic benefits in the agricultural field or in the medicine or veterinary medicine field may be obtained.

The ssRNA is usually administered as a pharmaceutical composition. The administration may be carried out by known methods, wherein a nucleic acid is introduced into a desired target cell in vitro or in vivo. Commonly used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation and microinjection and viral methods (Graham, F.L. and van der Eb, A.J. (1973) Virol. 52, 456; McCutchan, J.H. and Pagano, J.S. (1968), J. Natl. Cancer Inst. 41, 351; Chu, G. et al (1987), Nucl. Acids Res. 15, 1311; Fraley, R. et al. (1980), J. Biol. Chem. 255, 10431; Capecchi, M.R. (1980), Cell 22, 479). A recent addition to this arsenal of techniques for the introduction of nucleic acids into cells is the use of cationic liposomes (Felgner, P.L. et al. (1987), Proc. Natl. Acad. Sci USA 84, 7413). Commercially available cationic lipid formulations are e.g. Tfx 50 (Promega) or Lipofectamin2000 (Life Technologies). A further preferred method for the introduction of RNA into a target organism, particularly into a mouse, is the high-pressure tail vein injection (Lewis, D.L. et al. (2002), Nat.Genet.29, 29; McCaffrey, A.P. et al. (2002), Nature 418, 38-39).

Herein, a buffered solution comprising the single-stranded RNA (e.g. about 2 ml) is injected into the tail vein of the mouse within 10 s.

Thus, the invention also relates to a pharmaceutical composition containing as an active agent at least one single-stranded RNA molecule as described above and a pharmaceutical carrier. The composition may be used for diagnostic and for therapeutic applications in human medicine or in veterinary medicine.

For diagnostic or therapeutic applications, the composition may be in form of a solution, e.g. an injectable solution, a cream, ointment, tablet, suspension or the like. The composition may be administered in any suitable way, e.g. by injection, by oral, topical, nasal, rectal application etc. The carrier may be any suitable pharmaceutical carrier. Preferably, a carrier is used, which is capable of increasing the efficacy of the RNA molecules to enter the target-cells. Suitable examples of such carriers are liposomes, particularly cationic liposomes. A further preferred administration method is injection.

A further preferred application of the RNAi method is a functional analysis of eukaryotic cells, or eukaryotic non-human organisms, preferably mammalian cells or organisms and most preferably human cells, e.g. cell lines such as HeLa or 293 or rodents, e.g. rats and mice. By transfection with suitable single-stranded RNA molecules which are homologous to a predetermined target gene or DNA molecules encoding a suitable single-stranded RNA molecule a specific knockout phenotype can be obtained in a target cell, e.g. in cell culture or in a target organism. The presence of short single-stranded RNA molecules does not result in an interferon response from the host cell or host organism.

In an especially preferred embodiment, the RNA molecule is administered associated with biodegradable polymers, e.g. polypeptides, poly(d,l-lactic-

co-glycolic acid) (PLGA), polylysine or polylysine conjugates, e.g. polylysine-graft-imidazole acetic acid, or poly(beta-amino ester) or microparticles, such as microspheres, nanoparticles or nanospheres. More preferably the RNA molecule is covalently coupled to the polymer or microparticle, wherein the covalent coupling particularly is effected via the 3'-terminus of the RNA molecule.

Further, the invention relates to a pharmaceutical composition for inhibiting the expression of a target transcript by RNAi comprising as an active agent a single-stranded RNA molecule having a length from 14 - 50, preferably 15 - 29 nucleotides wherein at least the 14 - 20 5'most nucleotides are substantially complementary to said target transcript.

Furthermore, the invention relates to a method for the prevention or treatment of a disease associated with overexpression of at least one target gene comprising administering a subject in need thereof a single-stranded RNA molecule having a length from 14 - 50, preferably 15-29 nucleotides wherein at least the 14 - 20 5'most nucleotides are substantially complementary to a target transcript in an amount which is therapeutically effective for RNAi.

Still, a further subject matter of the invention is a eukaryotic cell or a eukaryotic non-human organism exhibiting a target gene-specific knockout phenotype comprising an at least partially deficient expression of at least one endogeneous target gene wherein said cell or organism is transfected with at least one single-stranded RNA molecule capable of inhibiting the expression of at least one endogeneous target gene. It should be noted that the present invention allows the simultaneous delivery of several antisense RNAs of different sequences, which are either cognate to a different or the same target gene.

Gene-specific knockout phenotypes of cells or non-human organisms, particularly of human cells or non-human mammals may be used in analytic procedures, e.g. in the functional and/or phenotypical analysis of complex physiological processes such as analysis of gene expression profiles and/or proteomes. For example, one may prepare the knock-out phenotypes of human genes in cultured cells which are assumed to be regulators of alternative splicing processes. Among these genes are particularly the members of the SR splicing factor family, e.g. ASF/SF2, SC35, SRp20, SRp40 or SRp55. Further, the effect of SR proteins on the mRNA profiles of predetermined alternatively spliced genes such as CD44 may be analysed. Preferably the analysis is carried out by high-throughput methods using oligonucleotide based chips.

Using RNAi based knockout technologies, the expression of an endogeneous target gene may be inhibited in a target cell or a target organism. The endogeneous gene may be complemented by an exogeneous target nucleic acid coding for the target protein or a variant or mutated form of the target protein, e.g. a gene or a cDNA, which may optionally be fused to a further nucleic acid sequence encoding a detectable peptide or polypeptide, e.g. an affinity tag, particularly a multiple affinity tag. Variants or mutated forms of the target gene differ from the endogeneous target gene in that they encode a gene product which differs from the endogeneous gene product on the amino acid level by substitutions, insertions and/or deletions of single or multiple amino acids. The variants or mutated forms may have the same biological activity as the endogeneous target gene. On the other hand, the variant or mutated target gene may also have a biological activity, which differs from the biological activity of the endogeneous target gene, e.g. a partially deleted activity, a completely deleted activity, an enhanced activity etc.

The complementation may be accomplished by coexpressing the polypeptide encoded by the exogeneous nucleic acid, e.g. a fusion protein com-

prising the target protein and the affinity tag and the double stranded RNA molecule for knocking out the endogeneous gene in the target cell. This coexpression may be accomplished by using a suitable expression vector expressing both the polypeptide encoded by the exogeneous nucleic acid, e.g. the tag-modified target protein and the single-stranded RNA molecule or alternatively by using a combination of expression vectors. Proteins and protein complexes which are synthesized de novo in the target cell will contain the exogeneous gene product, e.g. the modified fusion protein. In order to avoid suppression of the exogeneous gene product expression by the RNAi molecule, the nucleotide sequence encoding the exogeneous nucleic acid may be altered on the DNA level (with or without causing mutations on the amino acid level) in the part of the sequence which is homologous to the single-stranded RNA molecule. Alternatively, the endogeneous target gene may be complemented by corresponding nucleotide sequences from other species, e.g. from mouse.

Preferred applications for the cell or organism of the invention is the analysis of gene expression profiles and/or proteomes. In an especially preferred embodiment an analysis of a variant or mutant form of one or several target proteins is carried out, wherein said variant or mutant forms are reintroduced into the cell or organism by an exogeneous target nucleic acid as described above. The combination of knockout of an endogeneous gene and rescue by using mutated, e.g. partially deleted exogeneous target has advantages compared to the use of a knockout cell. Further, this method is particularly suitable for identifying functional domains of the target protein. In a further preferred embodiment a comparison, e.g. of gene expression profiles and/or proteomes and/or phenotypic characteristics of at least two cells or organisms is carried out. These organisms are selected from:

(i) a control cell or control organism without target gene inhibition,

(ii) a cell or organism with target gene inhibition and

(iii) a cell or organism with target gene inhibition plus target gene complementation by an exogeneous target nucleic acid.

The method and cell of the invention may also be used in a procedure for identifying and/or characterizing pharmacological agents, e.g. identifying new pharmacological agents from a collection of test substances and/or characterizing mechanisms of action and/or side effects of known pharmacological agents.

Thus, the present invention also relates to a system for identifying and/or characterizing pharmacological agents acting on at least one target protein comprising:

(a) a eukaryotic cell or a eukaryotic non-human organism capable of expressing at least one endogeneous target gene coding for said target protein,

(b) at least one single-stranded RNA molecule capable of inhibiting the expression of said at least one endogeneous target gene by RNAi and

(c) a test substance or a collection of test substances wherein pharmacological properties of said test substance or said collection are to be identified and/or characterized.

Further, the system as described above preferably comprises:

(d) at least one exogeneous target nucleic acid coding for the target protein or a variant or mutated form of the target protein wherein said exogeneous target nucleic acid differs from the endogeneous target gene on the nucleic acid level such that the expression of the exogeneous target nucleic acid is substantially less inhibited by the single-stranded RNA molecule than the expression of the endogeneous target gene.

Furthermore, the RNA knockout complementation method may be used for preparative purposes, e.g. for the affinity purification of proteins or protein complexes from eukaryotic cells, particularly mammalian cells and more particularly human cells. In this embodiment of the invention, the exogeneous target nucleic acid preferably codes for a target protein which is fused to an affinity tag.

The preparative method may be employed for the purification of high molecular weight protein complexes which preferably have a mass of ≥ 150 kD and more preferably of ≥ 500 kD and which optionally may contain nucleic acids such as RNA. Specific examples are the heterotrimeric protein complex consisting of the 20 kD, 60 kD and 90 kD proteins of the U4/U6 snRNP particle, the splicing factor SF3b from the 17S U2 snRNP consisting of 5 proteins having molecular weights of 14, 49, 120, 145 and 155 kD and the 25S U4/U6/U5 tri-snRNP particle containing the U4, U5 and U6 snRNA molecules and about 30 proteins, which has a molecular weight of about 1.7 MD.

This method is suitable for functional proteome analysis in mammalian cells, particularly human cells.

Finally, the invention relates to a purified and isolated mammalian, particularly human RNA-induced silencing complex (RISC) having an apparent molecular weight of less than about 150-160 kDa, e.g. about 120 to 150-160 kDa. The RISC comprises polypeptide and optionally nucleic acid components, particularly single-stranded RNA molecules as described above. The RISC may be used as a target for diagnosis and/or therapy, as a diagnostic and/or therapeutic agent itself, as a molecular-biological reagent or as component in a screening procedure for the identification and/or characterization of pharmaceutical agents.

Polypeptide components of RISC preferably comprise members of the Argonaute family of proteins, and contain eIF2C1 and/or eIF2C2, and possibly at least one other expressed eIF2C family member, particularly selected from eIF2C3, eIF2C4, HILI and HIWI.

Expression or overexpression of one or several proteins present in RISC in suitable host cells, e.g. eukaryotic cells, particularly mammalian cells, is useful to assist an RNAi response. These proteins may also be expressed or overexpressed in transgenic animals, e.g. vertebrates, particularly mammals, to produce animals particularly sensitive to injected single-stranded or double-stranded siRNAs. Further, the genes encoding the proteins may be administered for therapeutic purposes, e.g. by viral or non-viral gene delivery vectors.

It is also conceivable to administer a siRNA/eIF2C1 or 2 complex directly by the assistance of protein transfection reagents (e.g. Amphoteric Protein Transfection Reagents, ProVectin protein (Imgenex), or similar products) rather than RNA/DNA transfection. This may have technical advantages over siRNA transfection that are limited to nucleic acid transfection.

Alternatively to the application of siRNAs as synthetic double-stranded or single-stranded siRNAs, it is conceivable to also administer an antisense siRNA precursor molecule in the form of a hairpin stem-loop structure comprising 19 to 29 base pairs in the stem with or without 5' or 3' overhanging ends on one side of the duplex and a nucleotide or non-nucleotide loop on the other end. Preferably, the hairpin structure has a 3' overhang of from 1-5 nucleotides. Further, the precursor may contain modified nucleotides as described above, particularly in the loop and/or in the 3' portion, particularly in the overhang. The siRNA or precursors of siRNAs may also be introduced by viral vectors or RNA expression systems into a RISC compound, e.g. eIF2C1 and/or 2 overexpressing organism or cell line. The siRNA precursors may also be generated by direct expression

within an organism or cell line. This may be achieved by transformation with a suitable expression vector carrying a nucleic acid template operatively linked to an expression control sequence to express the siRNA precursor.

Further, the present invention is explained in more detail in the following figures and examples.

Figure legends

Figure 1. HeLa cytoplasmic S100 extracts show siRNA-dependent target RNA cleavage.
(A) Representation of the 177-nt ^{32}P-cap-labeled target RNA with the targeting siRNA duplex. Target RNA cleavage site and the length of the expected cleavage products is also shown. The fat black line positioned under the antisense siRNA is used in the following figures as symbol to indicate the region of the target RNA, which is complementary to the antisense siRNA sequence. (B) Comparison of the siRNA mediated target RNA cleavage using the previously established *D. melanogaster* embryo in vitro system and HeLa cell S100 cytoplasmic extract. 10 nM cap-labeled target RNA was incubated with 100 nM siRNA as described in materials. Reaction products were resolved on a 6% sequencing gel. Position markers were generated by partial RNase T1 digestion (T1) and partial alkaline hydrolysis (OH) of the cap-labeled target RNA. The arrow indicates the 5' cleavage product, the 3' fragment is unlabeled and therefore invisible.

Figure 2. Chemical modification of the 5' end of the antisense but not the sense siRNAs prevents sense target RNA cleavage in HeLa S100 extracts.
(A) Illustration of the possible 5' and 3' aminolinker modifications of the sense and antisense strands of a siRNA duplex. L5 represents a 6-carbon chain aminolinker connected via a 5'-phosphodiester linkage, L3 represents

a 7-carbon aminolinker connected via a phosphodiester bond to the terminal 3′ phosphate. s, sense; as, antisense. (B) Target RNA cleavage testing various combinations of 5′ and 3′ aminolinker-modified siRNA duplexes. NC (negative control) shows an incubation reaction of the target RNA in the absence of siRNA duplex. T1, RNase T1 ladder; OH, partial alkaline hydrolysis ladder.

Figure 3. siRNA containing 3′-terminal phosphates are subjected to ligation as well as dephosphorylation reactions.

(A) Sequence of the radiolabeled siRNA duplex. The labeled nucleotide was joined to synthetic 20-nt antisense siRNA by T4 RNA ligation of ^{32}pCp. The various combinations of 5′ and 3′ hydroxyl/phosphate were prepared as described in materials. X and Y indicate 5′ and 3′ modifications of the antisense siRNA. (B) Fate of the antisense siRNA during incubation of the modified siRNA duplexes in HeLa S100 extract in the presence of non-radiolabeled target RNA. The different phosphorylated forms of the antisense siRNA were distinguished based on their gel mobility. Identical results were obtained when using 5′ phosphorylated sense siRNA or when leaving out the target RNA during incubation. Ligation products are only observed when 3′ phosphates were present on the labeled antisense siRNA.

Figure 4. RISC is a stable complex that does not rapidly exchange bound siRNA.

Increasing concentrations of non-specific siRNA compete with target-specific RISC formation when added simultaneously to HeLa S100 extracts (lanes 4 to 7). However, when the unspecific siRNA duplex is added 15 min after pre-incubation with the specific siRNA duplex, no more competition was observed (3 lanes to the right). T1, RNase T1 ladder.

Figure 5. Partial purification of human RISC.

(A) Graphical representation of the structure of the biotinylated siRNA duplex used for affinity purification of siRNA-associated factors. L3 indicates a C7-aminolinker that was conjugated to a photo-cleavable biotin N-hydroxysuccinimidyl ester; UV indicates photocleavage of the UV-sensitive linkage to release affinity selected complexes under native conditions. (B) Superdex-200 gel filtration analysis of siRNA-protein complexes (siRNPs) recovered by UV treatment/elution (UV elu) from the streptavidin affinity column. Fractions were assayed for their ability to sequence-specifically cleave the cap-labeled target RNA. The number of the collected fractions and the relative positions of the aldolase (158 kDa) and BSA (66 kDa) size markers are indicated. (C) Glycerol gradient (5%-20%) sedimentation of siRNPs recovered by UV treatment/elution from the streptavidin affinity column. For legend, see (B). When monitoring the precise size of target RNA cleavage fragments using internally ^{32}P-UTP-labeled, capped mRNA, the sum is equal to the full-length transcript, thus indicating that target RNA is indeed only cleaved once in the middle of the region spanned by the siRNA.

Figure 6. RISC contains a single-stranded siRNA.
siRNPs were subjected to affinity selection after incubation using siRNA duplexes with one or both strands biotinylated. The eluate recovered after UV treatment or the unbound fraction after streptavidin affinity selection (flow-through) was assayed for target RNA degradation. If the antisense strand was biotinylated, all sense target RNA-cleaving RISC was bound to the streptavidin beads, while sense siRNA biotinylation resulted in RISC activity of the flow-through. The cleavage reaction in the flow-through fraction was less efficient than in the UV eluate, because affinity-selected RISC was more concentrated.

Figure 7. Single-stranded antisense siRNAs reconstitute RISC in HeLa S100 extracts.

Analysis of RISC reconstitution using single-stranded or duplex siRNAs comparing HeLa S100 extracts (A) and the previously described *D. melanogaster* embryo lysate (B). Different concentrations of single-stranded siRNAs (s, sense; as, antisense) and duplex siRNA (ds) were tested for specific targeting of cap-labeled substrate RNA. 100 nM concentrations of the antisense siRNA reconstituted RISC in HeLa S100 extract, although at reduced levels in comparison to the duplex siRNA. Reconstitution with single-stranded siRNAs was almost undetectable in *D. melanogaster* lysate, presumably because of the higher nuclease activity in this lysate causing rapid degradation of uncapped single-stranded RNAs .

Figure 8. Single-stranded antisense siRNAs mediate gene silencing in HeLa cells.

(A) Silencing of nuclear envelope protein lamin A/C. Fluorescence staining of cells transfected with lamin A/C-specific siRNAs and GL2 luciferase (control) siRNAs. Top row, staining with lamin A/C specific antibody; middle row, Hoechst staining of nuclear chromatin; bottom row, phase contrast images of fixed cells. (B) Quantification of lamin A/C knockdown after Western blot analysis. The blot was stripped after lamin A/C probing and reprobed with vimentin antibody. Quantification was performed using a Lumi-Imager (Roche) and LumiAnalyst software to quantitate the ECL signals (Amersham Biosciences), differences in gel loading were corrected relative to non-targeted vimentin protein levels. The levels of lamin A/C protein were normalized to the non-specific GL2 siRNA duplex.

Figure 9. Antisense siRNAs of different length direct target RNA cleavage in HeLa S100 extracts.

(A) Graphical representation of the experiment. Antisense siRNAs were extended towards the 5´side (series 1, 20 to 25-nt) or the 3´side (series 2, 20 to 23-nt).

(B) Target RNA cleavage using the antisense siRNAs described in (A). HeLa S100 extract was incubated with 10 nM cap-labeled target RNA and 100

nM antisense siRNAs at 30 °C for 2.5 h. Reaction products were resolved on a 6 % sequencing gel. Position markers were generated by partial RNase T1 digestion (T1) and partial alkaline hydrolysis (OH) of the cap-labeled target RNA. Arrows indicate the position of the 5´cleavage products generated by the different antisense siRNAs. The fat black lines on the left (series 1) and the right (series 2) indicate the region of the target RNA, which is complementary to the antisense siRNA sequences.

Figure 10. Length dependence of antisense siRNAs and effect of terminal modifications for targeting RNA cleavage in HeLa S100 extracts.
HeLa S100 extract was incubated with 10 nM cap-labeled target RNA and 100 nM antisense siRNAs at 30°C for 2.5h. Reaction products were resolved on a 6% sequencing gel. Position markers were generated by partial RNase T1 digestion (T1) of the cap-labeled target RNA. The fat black line on the left indicates the region of the target RNA, which is complementary to the 21-nt antisense siRNA sequence. The siRNA sequences used in each experiment are listed below (sense and antisense siRNAs are listed together, they were pre-annealed to form duplex siRNAs). p, phosphate; t, 2'-deoxythymidine, c, 2'-deoxycytidine, g, 2'-deoxycytidine, g, 2'-deoxyguanosine; L, aminolinker, B, photocleavable biotin; A,C,G,U, ribonucleotides.

Lane	Sense siRNA (5' – 3')	Antisense siRNA (5' – 3')
1		pUCGAAGUAUUCCG CG
2		pUCGAAGUAUUCCG CGUACGUG

3		pUCGAAGUAUUCCG CGUACGUGAUGU
4		pUCGAAGUAUUCCG CGUACGUGAUGUUC
5		pUCGAAGUAUUCCG CGUACGUGAUGUUC AC
6		pUCGAAGUAUUCCG CG
7		pUCGAAGUAUUCCG CGUACGUG
8		pUCGAAGUAUUCCG CGUACGUGAUGU
9		pUCGAAGUAUUCCG CGUACGUGAUGUUC
10		pUCGAAGUAUUCCG CGUACGUGAUGUUC AC
11		pUCGAAGUAUUCCG CGUACGUG
12		pUCGAAGUAUUCCG CGUACGtg
13		pUCGAAGUAUUCCG CGUACGUU
14		pUCGAAGUAUUCCG CGUACGtt

15		pUCGAAGUAUUCCG CGUACGUG
16		pUCGAAGUAUUCCG CGUACGtg
17		pUCGAAGUAUUCCG CGUACGUU
18		pUCGAAGUAUUCCG CGUACGtt
19	CGUACGCGGAAUACUUCG AAA	pUCGAAGUAUUCCG CGUACGUG
20	CGUACGCGGAAUACUUCG AAA	pUCGAAGUAUUCCG CGUACGtg
21	CGUACGCGGAAUACUUCG AAA	pUCGAAGUAUUCCG CGUACGUU
22	CGUACGCGGAAUACUUCG AAA	pUCGAAGUAUUCCG CGUACGtt
23		tCGAAGUAUUCCGC GUACGUULB
24	cGUACGCGGAAUACUUCG AUULB	tCGAAGUAUUCCGC GUACGUULB
25		ptCGAAGUAUUCCGC GUACGttLB
26	cGUACGCGGAAUACUUCG AttLB	ptCGAAGUAUUCCGC GUACGttLB

27		ptCGAAGUAUUCCGC GUACGttL

Figure 11: Single-stranded antisense siRNAs mediate gene silencing in HeLa cells.

Quantification of lamin A/C knockdown after Western blot analysis. The blot was stripped after lamin A/C probing and reprobed with vimentin antibody. Quantification was performed using a Lumi-Imager (Roche) and LumiAnalyst software to quantitate the ECL signals (Amersham Biosciences), differences in gel loading were corrected relative to non-targeted vimentin protein levels. The levels of lamin A/C protein were normalized to the non-specific GL2 siRNA duplex.

Figure 12. Protein composition of affinity purified RISC.

(A) Silver-stained SDS-PAGE gel of affinity-selected ribonucleoprotein complexes after glycerol gradient (5%-20%) sedimentation. The arrow indicates the band containing eIF2C1 and eIF2C2. Molecular size markers are indicated on the left. The asterisk indicates a fraction for which the protein pellet was lost after precipitation. (B) Target RNA cleavage assay of the collected fractions. RISC activity peaked in fraction 7 and 8; bu, buffer.

Figure 13. Mass spectrometric characterization of eIF2C1 and eIF2C2.

The 100 kDa band was analysed by mass spectrometry. Mass spectrum indicating the peptide peaks corresponding to eIF2C2 (A) and eIF2C1 (B). (C) Alignment of eIF2C2 and eIF2C1 amino-acid sequences indicating the position of the identified peptides. Sequence differences are indicated by yellow boxes.

Figure 14. Predicted amino-acid sequences of the six human Argonaute protein family members.

Figure 15. Alignment of the sequences of the six human Argonaute protein family members.
Predicted sequences of human eIF2C1-4, HILI and HIWI have been aligned using ClustalX program.

Figure 16. Predicted cDNA sequences of the six human Argonaute protein family members.

Figure 17. All members of the Argonaute family but HIWI are expressed in HeLa cells.
RT-PCR analysis on polyA RNA from HeLa cells. (A) Primers (forward and reverse) used for nested and semi-nested PCR amplification of the different Argonautes and expected length of the PCR products. (B) Agarose gel electrophoresis of the obtained PCR products, confirming the expected length. Left lanes, 100 bp DNA ladder.

Example

1. Material and methods

1.1 siRNA synthesis and biotin conjugation

siRNAs were chemically synthesized using RNA phosphoramidites (Proligo, Hamburg, Germany) and deprotected and gel-purified as described previously. 5' aminolinkers were introduced by coupling MMT-C6-aminolinker phosphoramidite (Proligo, Hamburg), 3' C7-aminolinkers were introduced by assembling the oligoribonucleotide chain on 3'-aminomodifier (TFA) C7 lcaa control pore glass support (Chemgenes, MA, USA). The sequences for GL2 luciferase siRNAs were as

described (Elbashir et al., 2001a, supra). If 5'-phosphates were to be introduced, 50 to 100 nmoles of synthetic siRNAs were treated with T4 polynucleotide kinase (300 µl reaction, 2.5 mM ATP, 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 30 U T4 PNK, New England Biolabs, 45 min, 37 °C) followed by ethanol precipitation.

3' Terminal ^{32}pCp labeling (Figure 3) was performed in a 30 µl reaction (17 µM siRNA, 0.5 µM ^{32}pCp (110 TBq/mmol), 15% DMSO, 20 U T4 RNA ligase, NEB, and 1x NEB-supplied reaction buffer) for 1.5 h at 37 °C, and gel-purified. One half of the pCp-labeled RNA was dephosphorylated (25 µl reaction, 500 U alkaline phosphatase, Roche, and Roche-supplied buffer, 30 min, 50 °C), followed by phenol/chloroform extraction and ethanol precipitation. Half of this reaction was 5' phosphorylated (20 µl reaction, 2 units T4 polynucleotide kinase, NEB, 10 mM ATP, NEB-supplied buffer, 60 min, 37 °C). A quarter of the initial pCp-labeled siRNA was also 5' phosphorylated (10 µl reaction, 10 units 3' phosphatase-free T4 polynucleotide kinase, Roche, 10 mM ATP, Roche-supplied buffer, 3 min, 37 °C).

For conjugation to biotin, 20 to 65 nmoles of fully deprotected aminolinker-modified siRNA were dissolved in 100 µl of 100 mM sodium borate buffer (pH 8.5) and mixed with a solution of 1 mg of EZ-Link NHS-PC-LC-Biotin (Pierce, IL, USA) in 100 µl of anhydrous dimethylformamide. The solution was incubated for 17 h at 25 °C in the dark. Subsequently, siRNAs were precipitated by the addition of 60 µl 2 M sodium acetate (pH 6.0) and 1 ml ethanol. The RNA pellet was collected by centrifugation and biotin-conjugated siRNA was separated from non-reacted siRNA on a preparative denaturing 18% acrylamide gel (40 cm length) in the dark. The RNA bands were visualized by 254 nm UV shadowing and minimized exposure time. The bands were excised, and the RNA was eluted overnight in 0.3 M NaCl at 4 °C and recovered by ethanol

precipitation. siRNA duplexes were formed as previously described (Elbashir et al., Methods 26 (2002), 199 - 213).

1.2 Preparation of S100 extracts from HeLa cells

Cytoplasm from HeLa cells adapted to grow at high density was prepared following the Dignam protocol for isolation of HeLa cell nuclei (Dignam et al., Nucleic Acids Res. 11 (1983), 1475 - 1489). The cytoplasmic fraction was supplemented with KCl, $MgCl_2$ and glycerol to final concentrations of 100 mM, 2 mM and 10%, respectively. At this stage, the extracts can be stored frozen at -70 °C after quick-freezing in liquid nitrogen without loss of activity. S100 extracts were prepared by ultracentrifugation at 31.500 rpm for 60 minutes at 4 °C using a Sorvall T-865 rotor. The protein concentration of HeLa S100 extract varied between 4 to 5 mg/ml as determined by Bradford assay.

1.3 Affinity purification of RISC with 3' biotinylated siRNA duplexes

For affinity purification of siRNA-associated protein complexes from HeLa S100 extracts, 10 nM of a 3' double-biotinylated siRNA duplex were incubated in 0.2 mM ATP, 0.04 mM GTP, 10 U/ml RNasin, 6 µg/ml creatine kinase, and 5 mM creatine phosphate in 60% S100 extract at 30 °C for 30 to 60 min and gentle rotation. Thereafter, 1 ml slurry of Immobilized Neutravidin Biotin Binding Protein (Pierce, IL, USA) was added per 50 ml of reaction solution and the incubation was continued for another 60 to 120 min at 30 °C with gentle rotation. The Neutravidin beads were then collected at 2000 rpm for 2 minutes at 4 °C in a Heraeus Megafuge 1.0 R centrifuge using a swinging bucket rotor type 2704. Effective capturing of RISC components after affinity selection was confirmed by assaying the supernatant for residual RISC activity with and without supplementing fresh siRNA duplexes. The collected Neutravidin beads were washed with 10 volumes of buffer A relative to the bead volume (30 mM HEPES, pH 7.4, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT, 10% glycerol) followed by washing with 5 volumes of buffer B (same as

buffer A with only 3% glycerol content). The beads were transferred to a 0.8 x 4 cm Poly-Prep chromatography column (BioRad; CA, USA) by resuspending in 3 volumes of buffer B at 4°C, followed by 10 volumes of washing with buffer B. Washing of the beads was continued by 10 volumes of buffer B increased to 300 mM KCl. The column was then reequilibrated with regular buffer B. To recover native siRNA-associated complexes, the column was irradiated in the cold room by placing it at a 2 cm distance surrounded by four 312 nm UV lamps (UV-B tube, 8 W, Herolab, Germany) for 30 minutes. To recover the photocleaved siRNP solution, the column was placed into a 50 ml Falcon tube and centrifuged at 2000 rpm for 1 minute at 4 °C using again the 2704 rotor. For full recovery of siRNPs, the beads were once again resuspended in buffer B followed by a second round of UV treatment for 15 minutes. Both eluates were pooled and assayed for target RNA degradation.

1.4 Target RNA cleavage assays

Cap-labeled target RNA of 177 nt was generated as described (Elbashir et al., EMBO J. 20 (2001 c), 6877 - 6888) except that his-tagged guanylyl transferase was expressed in *E. coli* from a plasmid generously provided by J. Wilusz and purified to homogeneity. If not otherwise indicated, 5' phosphorylated siRNA or siRNA duplex was pre-incubated in supplemented HeLa S100 extract at 30°C for 15 min prior to addition of cap-labeled target RNA. After addition of all components, final concentrations were 100 nM siRNA, 10 nM target RNA, 1 mM ATP, 0.2 mM GTP, 10 U/ml RNasin, 30 µg/ml creatine kinase, 25 mM creatine phosphate, 50% S100 extract. Incubation was continued for 2.5 h. siRNA-mediated target RNA cleavage in *D. melanogaster* embryo lysate was performed as described (Zamore et al., Cell 101 (2000), 25 - 33). Affinity-purified RISC in buffer B was assayed for target RNA cleavage without preincubation nor addition of extra siRNA (10 nM target RNA, 1 mM ATP, 0.2 mM GTP, 10 U/ml RNasin, 30 µg/ml creatine kinase, 25 mM creatine phosphate, 50% RISC in buffer B). Cleavage reactions were stopped by the addition of 8 vols of

proteinase K buffer (200 mM Tris–HCl pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% w/v SDS). Proteinase K, dissolved in 50 mM Tris-HCl pH 8.0, 5 mM $CaCl_2$, 50% glycerol, was added to a final concentration of 0.6 mg/ml and processed as described (Zamore et al. (2000), supra). Samples were separated on 6% sequencing gels.

1.5 Analytical Gel Filtration

UV-eluates in buffer B were fractionated by gel filtration using a Superdex 200 PC 3.2/30 column (Amersham Biosciences) equilibrated with buffer A on a SMART system (Amersham Biosciences). Fractionation was performed by using a flow rate of 40 µl/minute and collecting 100 µl fractions. Fractions were assayed for specific target RNA cleavage. Size calibration was performed using molecular size markers thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa) and BSA (66 kDa) (Amersham Biosciences).

1.6 Glycerol gradient sedimentation

UV-eluates were layered on top of 4 ml linear 5% to 20% (w/w) glycerol gradient adjusted to 30 mM HEPES, pH 7.4, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT. Centrifugation was performed at 35000 rpm for 14.5 h at 4 °C using a Sorvall SW 60 rotor. Twenty fractions of 0.2 ml volume were removed sequentially from the top and 15 µl aliquots were used to assay for target RNA cleavage.

2. Results

2.1 A human biochemical system for siRNA functional analysis

We were interested in assaying siRNA-mediated target RNA degradation in human cell extracts, because siRNAs are powerful reagents to knockdown gene expression in human cells but the action of siRNAs in human cells was uncertain. To investigate whether siRNAs guide target RNA

degradation in human cells with a similar mechanism to the one observed in *D. melanogaster* (e.g. Elbashir et al. (2001 b), supra), we prepared substrates for targeted mRNA degradation as described previously (Elbashir et al. (2001 c), supra). A 5'-^{32}P-cap-labeled, 177-nt RNA transcript, derived from a segment of the firefly luciferase gene, was incubated in HeLa cell S100 or *D. melanogaster* embryo extracts with a 21-nt siRNA duplex in the presence of an ATP regeneration system (Figure 1A, B). siRNA cleavage assays were performed at 25 °C in *D. melanogaster* lysate and at 30 °C in HeLa S100 extracts for 2.5 h. After deproteinization using proteinase K, the reaction products were separated on a 6% sequencing gel.

Similar to the previous observation in *D. melanogaster* lysate, we observed the appearance of a cleavage product in HeLa S100 extract at exactly the same position, thus indicating that the siRNA duplex guides target RNA cleavage in the human system with the same specificity and mechanism. The cleavage reaction appeared less efficient when compared to the *D. melanogaster* system, but this could be explained by the 5-fold lower total protein concentration of HeLa S100 extracts (25 mg/ml vs. 5 mg/ml). Similar to *D. melanogaster* lysates, siRNA duplexes without 5' phosphate were rapidly 5' phosphorylated in HeLa S100 extracts (see below) and the ability to cleave the target RNA was independent of the presence of a 5' phosphate on the synthetic siRNA duplexes.

Comparative analysis of the efficiency of siRNA duplexes of different length in *D. melanogaster* lysate and in transfected mammalian cells indicated that the differences in silencing efficiencies between 20- to 25-nt siRNA duplexes were less pronounced in mammalian cells than in *D. melanogaster* (Elbashir et al. (2002), supra) . Duplexes of 24- and 25-nt siRNAs were inactive in *D. melanogaster* lysate, whereas the same duplexes were quite effective for silencing when introduced by transfection into HeLa cells. We therefore asked whether siRNA duplexes of 20- to 25-nt are able to reconstitute RISC also with approximately equal

efficiency. Indeed, we observed no large differences in our biochemical assay, and the position of target RNA cleavage was as predicted according to the cleavage guiding rules established in *D. melanogaster* lysate (data not shown). Our biochemical results therefore support the in vivo observations.

2.2 5' modification of the guide siRNA inhibits RISC activity

Modification of siRNAs at their termini is important for developing siRNA-based affinity purification schemes or for conjugating reporter tags for biophysical measurements. The most common method for introducing reactive side chains into nucleic acids is by chemical synthesis using aminolinker derivatives (Eckstein (1991), Oligonucleotides and analogues, 2nd Ed., Oxford UK, Oxford University Press). After complete deprotection of the oligonucleotide, the primary amine is typically reacted with the N-hydroxysuccinimidyl ester of the desired compound. We have introduced 5' and 3' aminolinkers with six and seven methylene groups as spacers, respectively. The linker-modified siRNA duplexes were tested for mediating target RNA degradation in HeLa S100 extract (Figure 2A, B). Modification of the 5'- end of the antisense guide siRNA abolished target RNA cleavage, while modification of neither the sense 5'-end nor of both 3'-ends showed any inhibitory effect. In an identical experiment using *D. melanogaster* embryo lysate, we observed a similar pattern of RISC activity although the duplex carrying the 5' aminolinker-modified antisense siRNA showed some residual activity (data not shown). Presumably, introduction of additional atoms or the change in terminal phosphate electric charge at the 5'-end of the antisense siRNA interfered with its ability to function as guide RNA. The critical function of the guide siRNAs 5' end was previously documented (Elbashir et al. (2001 c), supra).

The ability to modify siRNAs at their 3'-end suggests that siRNAs do not play a major role for priming dsRNA synthesis and do not act as primers for degenerative PCR. The fate of a siRNA in HeLa S100 extracts was

followed directly by incubation of an internally ^{32}pCp-radiolabeled siRNA duplexes. The radiolabeled antisense siRNA strand was also prepared with different 5′ and 3′ phosphate modifications (Figure 3A). All described combinations of siRNA duplexes were fully competent for RISC-dependent target RNA degradation (data not shown). As previously observed for *D. melanogaster* lysates (Nykänen et al. (2001), supra), rapid 5′ phosphorylation of siRNA duplexes with free 5′ hydroxyl termini was apparent. To our surprise, we noted that a small fraction of the 3′ phosphorylated antisense siRNA could be ligated to the opposing 5′ hydroxyl of the sense siRNA producing a lower mobility band. The inter-strand ligation was confirmed by changing the length of the unlabeled sense siRNA, which resulted in the expected mobility changes of the ligation product (data not shown). RNA ligase activity was previously observed in HeLa S100 extracts and it is mediated by two enzymatic activities (e.g. Vicente and Filipowicz, Eur. J. Biochem., 176 (1988), 431 - 439). The 3′ terminal phosphate is first converted to a 2′,3′-cyclic phosphate requiring ATP and 3′ terminal phosphate cyclase. Thereafter, the opposing 5′ hydroxyl is ligated to the cyclic phosphate end by an as yet uncharacterized RNA ligase. We chemically synthesized the predicted 5′ phosphorylated, 42-nt ligation product and found that it is unable to mediate target RNA cleavage, presumably because it can not form activated RISC. The majority of the 3′ phosphorylated duplexes siRNA was gradually dephosphorylated at its 3′ end and emerged chemically similar to naturally generated siRNA. Together, these observations indicate that the cell has a mechanism to preserve the integrity of siRNAs. We were unable to detect a proposed siRNA-primed polymerization product (Figure 3B), suggesting that siRNAs do not function as primers for template-dependent dsRNA synthesis in our system. However, we acknowledge that a proposed RNA-dependent polymerase activity may have been inactivated during preparation of our extracts.

2.3 siRNAs incorporated into RISC do not compete with a pool of free siRNAs

In order to analyze RISC assembly and stability, we tested whether target-unspecific siRNA duplexes were able to compete with target-specific siRNA duplexes. When specific and non-specific siRNA duplexes were co-incubated in HeLa S100 extracts, increasing concentrations of unspecific siRNA duplex competed with the formation of target-specific RISC (Figure 4, left lanes). However, when target-specific siRNAs were pre-incubated in HeLa S100 extract for 15 min in the absence of competitor siRNA duplex, the assembled siRNA in the target-specific RISC could no longer be competed with the target-unspecific siRNA duplex (Figure 4, right lanes). This result suggests that RISC is formed during the first 15 minutes of incubation and that siRNAs were irreversibly associated with the protein components of RISC during the 2.5 h time window of the experiment.

2.4 Purification of human RISC

After having the 3' termini of siRNAs defined as the most suitable position for chemical modification, a photo-cleavable biotin derivative was conjugated to the 3' aminolinker-modified siRNAs. A photo-cleavable biotin derivative was selected because of the advantage of recovering RISC under non-denaturing conditions after capturing complexes on streptavidin-coated affinity supports. 3' Conjugation of biotin to the sense, antisense or to both of the strands did not affect target RNA cleavage when compared to non-biotinylated siRNAs (data not shown). siRNA duplexes with biotin residues on both 3' ends were therefore used for affinity purification (Figure 5A). The double biotinylated siRNA duplex was incubated in HeLa S100 extracts in the presence of ATP, GTP, creatine phosphate, and creatine kinase for ATP regeneration. Thereafter, streptavidin-conjugated agarose beads were added to capture the biotinylated siRNA

ribonucleoprotein complexes (siRNPs) including RISC. After extensive washing of the collected beads, the siRNPs were released by UV irradiation at 312 nm. The eluate cleaved target RNA sequence-specifically, thus indicating that RISC was recovered in its native state from the resin (Figure 5B, C, lane UV elu). The flow-through from the affinity selection showed no detectable RISC activity indicating complete binding of RISC by the beads (Figure 6). The affinity eluate was further analyzed by applying it onto a Superdex 200 gel filtration column (Figure 5B) as well as a 5%-20% glycerol gradient ultra-centrifugation (Figure 5C). Individual fractions were collected and assayed for target RNA cleavage without the addition of any further siRNA. RISC activity appeared between the molecular size markers aldolase (158 kDa) and BSA (66 kDa) after gel filtration or glycerol gradient centrifugation (Figure 5B, C). The molecular size of human RISC is therefore estimated to be between 90 and 160 kDa, significantly smaller than the complex previously analyzed in *D. melanogaster* lysates (Hammond et al. (2000), supra; Nykänen et al. (2001), supra). The small size of RISC suggests that Dicer (210 kDa) is not contained in RISC and that the formation of RISC from synthetic siRNAs may occur independently of Dicer. While these results do not rule out a role for Dicer during assembly of RISC, they emphasize the absence of Dicer in RISC.

2.5 RISC contains a single siRNA strand and can be reconstituted using single-stranded siRNAs

Two models are currently discussed concerning the siRNA strand composition of RISC. The first model suggests that both strands of the initially added siRNA duplex are physically present in RISC, but in an unwound conformation. The second model proposes that RISC carries only a single siRNA strand, implying loss of one of the siRNA strands during assembly. The latter model has been favored based on the analogy to miRNA precursor processing, where only one 21-nt strand accumulated from a dsRNA hairpin precursor. The molecular basis for the asymmetry of

the miRNA precursor processing reaction is not yet understood. Because siRNAs have symmetric 2-nt 3'-overhangs it is assumed that siRNA duplexes enter RISC with equal probability for both orientations, thus giving rise to distinct sense and antisense targeting RISCs.

To address the constitution of siRNAs in RISC, we affinity selected the assembled complexes with siRNA duplexes that were biotinylated at only one of the two constituting strands or both (Figure 6). If both strands were present together in RISC, the cleavage activity should be affinity selected on Neutravidin independently of the position of the biotin residue. In contrast, we observed target RNA cleavage from UV eluates after streptavidin selection only for siRNA duplexes with biotin conjugated to the antisense strand, but not the sense strand (Figure 6). RISC activity, assembled on siRNA duplexes with only the sense siRNA biotinylated, remained in the flow-through. These data suggest that RISC contains only a single-stranded RNA molecule.

To assess whether single-stranded siRNAs may be able to reconstitute RISC, single-stranded 5' phosphorylated siRNAs as well as the siRNA duplex were incubated at concentrations between 1 to 100 nM with cap-labeled target RNA in HeLa S100 extract (Figure 7A). At 100 nM single-stranded antisense siRNA, we detected RISC-specific target RNA cleavage, thus confirming that single-stranded siRNAs are present in RISC. At lower concentrations of single-stranded siRNAs, RISC formation remained undetectable while duplex siRNAs were effectively forming RISC even at 1 nM concentration. Therefore, a specific pathway exists which converts double-stranded siRNA into single-stranded siRNA containing RISC. Using *D. melanogaster* embryo lysate, we were unable to detect RISC activity from antisense siRNA (Figure 7B), presumably because of the high load of single-strand specific ribonucleases (Elbashir et al. (2001 b), supra). Furthermore, 5' phosphorylated 20- to 25-nt antisense siRNAs were able to mediate RISC-specific target RNA degradation in HeLa S100

extract producing the same target RNA cleavage sites as duplex siRNAs of this length (data not shown).

Finally, we tested single-stranded and duplex siRNAs for targeting of an endogenous gene in HeLa cells following our standard protocol previously established for silencing of lamin A/C. 200 nM concentrations of single-stranded siRNAs with and without 5' phosphate and 100 nM concentrations of duplex siRNAs were transfected into HeLa cells. Lamin A/C levels were monitored 48 h later using immunofluorescence (Figure 8A) and quantitative luminescence-based Western blot analysis (Figure 8B). Non-phosphorylated antisense siRNA caused a substantial knockdown of lamin A/C to about 25% of its normal level while 5' phosphoryled siRNAs reduced the lamin A/C content to less than 5%, similar to the reduction observed with the lamin A/C 5' phosphorylated (data not shown) or non-phosphorylated duplex siRNA (Figure 8). Sense siRNA and GL2 unspecific siRNA did not affect lamin A/C levels. The levels of non-targeted vimentin protein were monitored and used for normalizing of the loading of the lanes of the lamin A/C Western blots.

Gene silencing was also observed with phosphorylated as well as non-phosphorylated antisense siRNAs ranging in size between 19 to 29 nt. The phosphorylated antisense siRNAs were consistently better performing than the non-phosphorylated antisense, and their silencing efficiencies were comparable to that of the conventional duplex siRNA (Figure 11).

2.6 Protein composition of RISC

In order to identify the protein components of the RNA-induced silencing complex (RISC) in HeLa S100 extract, the specific affinity selection previously outlined was used. UV eluates were fractionated on a 5-20% glycerol gradient, fractions were recovered from the gradient and analysed for protein composition and target RNA endonucleolytic activity.

Two proteins of approximately 100 kDa were identified by mass spectrometry in the peak fraction of the endonucleolytic activity (Figure 12, fractions 7 and 8), corresponding to eIF2C1 and eIF2C2/GERp95 (Figure 13A and B). These proteins are 82% similar and are both members of the Argonaute family (Figure 13C). The first evidence that Argonaute proteins are part of RISC was provided by classical biochemical fractionation studies using dsRNA-transfected D. melanogaster S2 cells (Hammond et al., 2001, supra). The closest relative to D. melanogaster Argonaute2, D. melanogaster Argonaute1, was recently shown to be required for RNAi (Williams and Rubin, PNAS USA 99 (2002), 6889-6894).

Mass spectrometry analysis also revealed the presence of three peptides belonging exclusively to the HILI member of the Argonaute family of proteins. The sequences of those peptides are: NKQDFMDLSICTR, corresponding to positions 17-29 of the protein; TEYVAESFLNCLRR, corresponding to positions 436-449 of the protein, and; YNHDLPARIIVYR, corresponding to positions 591-603 of the protein. This finding suggests that the protein HILI may also be part of RISC.

In human, the Argonaute family is composed of 6 members, eIF2C1, eIF2C2, eIF2C3, eIF2C4, HILI and HIWI (Figure 14). The alignment of the six predicted amino-acid sequences show a high conservation, in particular between the eIF2C members, and HILI and HIWI (Figure 15). Predicted cDNA sequences encoding the Argonaute proteins are also shown (Figure 16).

The expression of the human Argonaute proteins was also investigated in HeLa cells by RT-PCR analysis using total and poly (A) selected RNA. All members of the family but HIWI were detected (Figure 17).

3. Discussion

The development of a human biochemical system for analysis of the mechanism of RNAi is important given the recent success of siRNA duplexes for silencing genes expressed in human cultured cells and the potential for becoming a sequence-specific therapeutic agent. Biochemical systems are useful for defining the individual steps of the RNAi process and for evaluating the constitution and molecular requirements of the participating macromolecular complexes. For the analysis of RNAi, several systems were developed, with the *D. melanogaster* systems being the most comprehensive as they enable to reconstitute dsRNA processing as well as the mRNA targeting. For mammalian systems, reconstitution of the mRNA targeting reaction has not yet been accomplished. Here, we describe the development and application of a biochemical system prepared from the cytoplasmic fraction of human HeLa cells, which is able to reconstitute the human mRNA-targeting RNA-induced silencing complex (RISC). Formation of RISC was accomplished using either 5′ phosphorylated or non-phosphorylated siRNA duplexes; as well as single-stranded antisense siRNAs; non-phosphorylated siRNA duplexes and presumably also single-stranded antisense siRNAs are rapidly 5′ phosphorylated in HeLa cell extracts (Figure 3).

Biochemical characterization of siRNA function

Reconstitution of RISC activity was only observed using cytoplasmic HeLa extracts. HeLa nuclear extracts assayed under the same conditions did not support siRNA-specific target RNA cleavage, thus suggesting that RISC components are located predominantly in the cytoplasm (data not shown).

Modifications of the 5′ and 3′ termini of siRNAs were tested in order to assess the importance of the siRNA termini for the targeting step. It was found that the 5′ end modification of the guide siRNA was more inhibitory for target RNA cleavage than 3′ end modification. Introduction of the 3′

biotin affinity tag into the target-complementary guide siRNA enabled us to affinity select sense-RNA-targeting RISC, whereas 3' biotinylation of the sense siRNA strand resulted in RISC activity in the flowthrough. Furthermore, the single RNA strand composition of RISC was confirmed by reconstituting the sequence-specific endonuclease complex using 5'-phosphorylated single-stranded guide siRNA. The reconstitution of RISC from single-stranded siRNA was however less effective and required 10- to 100-fold higher concentrations compared to duplex siRNA. Reconstitution of RISC from single-stranded siRNA was undetectable using *D. melanogaster* embryo lysate, which is most likely explained by the high content of 5' to 3' exonucleases in embryo lysate.

The size of RISC in HeLa lysate was determined by gel filtration as well as glycerol gradient ultracentrifugation after streptavidin affinity purification with 3' biotinylated siRNA duplexes. Sizes for RISC in *D. melanogaster* systems have been reported within a range of less than 230 to 500 kDa , however size determinations were conducted without having affinity purified RISC. Our affinity-purified RISC sediments in a narrow range between the size makers of 66 and 158 kDa. The differences to the reported sizes for RISC are not species-specific as we observed a similar size for RISC in *D. melanogaster* S2 cell cytoplasmic extracts after affinity purification (data not shown).

It has previously been proposed that siRNAs act as primers for target RNA-templated dsRNA synthesis (Lipardi et al., Cell 107 (2001), 297 - 307) although homologs for such RNA-dependent RNA polymerases known to participate in gene silencing in other systems are not identified in *D. melanogaster* or mammalian genomes. Analysis of the fate of siRNA duplexes in the HeLa cell system did not provide evidence for such a siRNA-primed activity (Figure 3), but indicates that the predominant pathway for siRNA-mediated gene silencing is sequence-specific endonucleolytic target RNA degradation.

Single-stranded 5' phosphorylated antisense siRNAs as triggers of mammalian gene silencing

It was previously noted that introduction of sense and antisense RNAs of several hundred nucleotides in length into *C. elegans* was able to sequence-specifically silence homologous genes (Guo and Kemphues, Cell 81 (1995), 611 - 620). Later, it was suggested that the sense and antisense RNA preparation were contaminated with a small amount of dsRNA, which was responsible for the silencing effect and is a much more potent inducer of gene silencing (Fire et al. (1998), supra). It is however conceivable that antisense RNA directly contributed to initiation of silencing. Indeed, most recently it was shown that antisense RNAs between 22 and 40 nt, but not sense RNAs were able to activate gene silencing in *C. elegans* (Tijsterman et al., Science 295 (2002), 694 - 697). The authors, however, favored the hypothesis of siRNA-primed dsRNA synthesis.

We have shown that modification of the 3' ends of antisense siRNA did not interfere with reconstitution of RISC in the human system. Together, these observations suggest that the driving forces for gene silencing in C. elegans may be predominantly dsRNA synthesis followed by Dicer cleavage, while in human and possibly also in D. melanogaster RISC-specific target mRNA degradation predominates.

Targeting of endogenously expressed lamin A/C by transfection of duplex siRNA into HeLa cells was the first reported example of siRNA-induced gene silencing. Lamin A/C protein was drastically reduced by a lamin A/C-specific siRNA duplex within two days post transfection, while unspecific siRNA duplexes showed no effect. At the time, transfection of non-phosphorylated sense or antisense siRNA did not reveal a substantial effect on lamin A/C levels, although more recently a minor reduction upon antisense siRNA transfection was noticed when similar concentrations of antisense siRNA were delivered as described in this study. However, the

effect was not interpreted as RISC-specific effect. Assaying 5'-phosphorylated antisense siRNAs revealed a substantial increase in lamin A/C silencing. Probably, 5' phosphorylated siRNAs are more stable or enter RISC more rapidly. Alternatively, the 5' end of transfected single-stranded siRNA may be less rapidly phosphorylated in the cell in comparison to duplex siRNAs.

Finally, it should be noted that HeLa cells are generally poor in nucleases and represent one of the preferred mammalian systems for studying RNA-processing or transcription reactions in vivo and in vitro. However, it can be expected that 5' phosphorylated single-stranded antisense siRNAs are suitable to knockdown gene expression in other cell types or tissues with a different content of nucleases, since chemical strategies to improve nuclease resistance of single stranded RNA are available. The general silencing ability of various cell types may also depend on the relative levels of siRNA/miRNA-free eIF2C1 and eIF2C2 proteins capable of associating with exogenously delivered siRNAs.

In summary, single-stranded 5'-phosphorylated antisense siRNAs of 19- to 29-nt in size broaden the use of RNA molecules for gene silencing because they can enter the mammalian RNAi pathway in vitro as well as in vivo through reconstitution of RISC. Human eIF2C1 and/or eIF2C2 seem to play a critical role in this process. Considering the feasibility of modulating the stability and uptake properties of single-stranded RNAs, 5'-phosphorylated single-stranded antisense siRNAs may further expand the utility of RNAi-based gene silencing technology as tool for functional genomics as well as therapeutic applications.

Argonaute proteins are a distinct class of proteins, containing a PAZ and Piwi domain (Cerutti et al., 2000, supra) and have been implicated in many processes previously linked to post-transcriptional silencing, however only limited biochemical information is available.

Human eIF2C2 is the ortholog of rat GERp95, which was identified as a component of the Golgi complex or the endoplasmic reticulum and copurified with intracellular membranes (Cikaluk et al., Mol.Biol.Cell 10 (1999), 3357-3722). More recently, HeLa cell eIF2C2 was shown to be associated with microRNAs and components of the SMN complex, a regulator of ribonucleoprotein assembly, suggesting that eIF2C2 plays a role in miRNA precursor processing or miRNA function (Mourelatos et al., Genes & Dev.16 (2002), 720-728). A more provocative hypothesis is that miRNAs are also in a RISC-like complex, which could potentially mediate target RNA degradation, if only perfectly matched miRNA target mRNAs existed. Sequence analysis using cloned human and mouse, however, did not reveal the presence of such perfectly complementary sequences in the genomes (Lagos-Quintana et al., Science 294 (2001), 853-858). Therefore, miRNPs may only function as translational regulators of partially mismatched target mRNAs, probably by recruiting additional factors that prevent dissociation from mismatched target mRNAs.

Human eIF2C1 has not been linked to gene silencing previously, but it is more than 80% similar in sequence to eIF2C2 (Koesters et al., Genomics 61 (1999), 210-218). This similarity may indicate functional redundancy, but it is also conceivable that functional RISC may contain eIF2C1 and eIF2C2 heterodimers. The predicted molecular weight of this heterodimeric complex would be slightly larger than the observed size of 90-160 kDa, but because size fractionation is based on globular shape, we can not disregard this possibility at this time.

Due to the high conservation between the members of the Argonaute family, it is possible that peptides that derive from regions 100% conserved in the 6 predicted proteins, may belong to members others than eIF2C1 and eIF2C2. In this respect, three peptides were identified with masses corresponding to HILI, meaning that this protein might be also a component of RISC.

To precisely assess the protein composition of RISC, reconstitution of the siRNA-mediated target RNA cleavage must be achieved by using recombinant proteins which may be obtained by cloning and expression in suitable bacterial or eukaryotic systems.

We expect that the biochemical characterization or the siRNA-mediated target RNA degradation process will have immediate applications, such as the development of cell lines or transgenic animals overexpressing RISC components. The efficiency in targeting endogenous genes in those lines or organisms will be enhanced. Furthermore, a reconstituted in vitro system for RNAi will allow the design of more potent and specific siRNA to achieve gene silencing.